SWAC Estimating Procedure

Introduction

The Record of Decisions, for the Lower Fox River and Green Bay Superfund Sites, require remediation of all contaminated sediment exceeding the 1.0 ppm PCB Remedial Action Level (RAL) in OU1, OU 2 (Deposit DD), OU 3, OU 4, and OU5 (River Mouth) either by the primary remedial approach or by one of the alternate remedial approaches discussed in the applicable Record of Decision (ROD). Each ROD establishes two standards that will be used to judge the completion of construction of the Remedy in each Operating Unit (OU): a RAL Performance Standard and a Surface Weighted Average Concentration (SWAC) goal.

Construction of the remedy in an OU will be deemed complete if the RAL Performance Standard has been met throughout the OU. If the RAL Performance Standard has not been met after employing the primary remedial approach and/or the alternate remedial approaches throughout the OU, then the remedy will be deemed complete if the SWAC, as determined by WDNR and USEPA, meets the SWAC goal for an OU. The construction of the remedy will not be deemed complete based on the SWAC goal unless and until all sediment exceeding the RAL has been remediated using the primary remedial approach and/or the alternate remedial approaches.

The current intention of the WDNR and USEPA is to utilize the SWAC estimating procedure as presented herein. However, as more information is collected and field experience gained for these remedial projects, this SWAC estimating procedure could be modified at the discretion of the WDNR and USEPA.

Procedure

Regulatory decision documents associated with the Fox River PCB Superfund Site require that the surface weighted average concentration (SWAC) of PCBs within each operable unit (OU) achieve certain targets after completion of planned remedial activities. However, no documents have rigorously defined the term nor have statistically valid computational procedures been described for estimating this quantity. The objective of this report is to propose a rigorous definition of the SWAC as well as to provide statistically valid estimation methods including procedures to quantify uncertainty.

The SWAC could be estimated using a variety of sampling designs and corresponding analysis methods. This estimation procedure was motivated by the guiding principles to 1) develop an unbiased estimator, 2) develop an analysis method that would not require substantial additional field sampling beyond the certification data already proposed, 3) avoid model based estimators in order to minimize assumptions, and 4) develop a method for which uncertainty could be easily quantified. These principles lead to a design based approach that is common in environmental and ecological studies based on stratified random sampling designs.

It is anticipated that attainment of goals associated with SWAC will be based on these proposed methods and that uncertainty in estimates will be acknowledged and incorporated into the decision process.

The purpose for using SWAC as a measure of remedial success is motivated from the notion that risk to resources within aquatic systems is proportional to exposure to PCBs. Further it is thought that exposure is proportional to the concentrations within the biologically active layer of sediment. The thickness of the biologically active layer has not been conclusively defined for all species and process, but has often been referred to as the top 2 to 12 inches of sediment. For purposes of this document it is assumed that the responsible parties and agencies will agree to a fixed depth representative of "surface" concentrations appropriate for quantifying exposure and subsequently risk. The important aspect is that if surface sediments are defined to be the top 6 inches of sediment, data used to estimate SWAC must be representative of the top 6 inches of sediment. The depth of sediment samples should ideally coincide or be strongly associated with the defined thickness of surface sediments.

SWAC Definition:

SWAC is the ratio of total PCB mass to total sediment mass on a dry weight basis within the surface sediments of a pre-specified area of interest. This can be restated as the average dry weight PCB concentration within the surface sediments of the pre-specified area of interest.

Estimation:

Because PCB and sediment mass are only known from an incomplete sample of the target population of interest, it is necessary to use statistics to estimate the true population parameter and to quantify the uncertainty in the estimate. Deterministic calculations can and have been used to estimate the population SWAC, however these methods are of limited value due to the failure to quantify uncertainty due to sampling error as well as the potential biases associated with deterministic models that require subjective modeling choices. The methods proposed in this document are unbiased to the population parameters and provide methods to describe uncertainty due to statistical sampling. Other potential uncertainties due to particular data handling techniques are also incorporated.

Error:

Uncertainty in the estimated SWAC can be broadly partitioned into components associated with sampling variation and bias due to certain assumptions necessary to fill data gaps or to accommodate negotiated agreements between the companies and agencies.

Sampling Variation:

Because the SWAC is estimated with sample data there is uncertainty in the estimate that can be attributed to chance errors due to sampling. This type of error can be made arbitrarily small by increasing the number of samples. In the extreme situation, if all of the surface sediment was removed and the PCBs separated from the remaining material and weighed, the sampling error would be reduced to zero. In spite of highly non-normally distributed PCB concentrations, for large sample sizes used to estimate SWAC within operable units, sampling variation of the SWAC can be expected to be approximately normally distributed. Confidence intervals will be used to quantify uncertainty due to sampling variation.

Bias:

Areas that have not been sampled may require imputation of values based on professional judgment and previous experience with other similar areas of the site or other sites. Failure to correctly "guess" concentrations in these areas may result in a bias in the overall estimated SWAC. Bias can be reduced through additional studies and sampling in areas that have not been previously investigated. The potential effects of bias will be quantified by considering a range of plausible situations. In general, the SWAC estimate and its' confidence interval shifts with varying bias.

Stratified Sampling Design:

The SWAC estimation method described in this document is a design based estimator. Design based estimation procedures are directly linked to and determined by the sampling design. In this case sample data will be collected from a series of strata defined by varying treatment techniques. For example, all areas which are un-treated would define one stratum; areas that are sand covered would define another stratum and so forth. For the Fox River it is anticipated that there will be strata corresponding to:

1) no action (i.e., areas with soft sediment less than the RAL);
2) void areas (i.e., areas where sampling occurred but no soft sediments was recovered)
3) sand cover;
4) dredge only;
5) dredge and sand cover;
6) dredge and cap;
7) cap only; and
8) unsampled areas (no-action areas that have not been sampled).

The methods defined in this document are general and can accommodate any number of strata as needed.

It is assumed that sample data are collected within each stratum based on an un-biased sampling design. Qualifying sampling designs could include systematic grids or randomized designs. Sampling designs may vary among strata. For example one may implement a systematic design within the sand covered area and a simple random sampling design within the capped area. To account for varying designs and sample sizes, data are aggregated within strata and then combined appropriately across strata using standard stratified sampling formulas (Cochran 1977).

Definitions:

Suppose that there are $h=1,2,3\ldots,L$ distinct strata that have been sampled. Assume that each stratum has area A_h and that the total area given by the sum of the stratum areas is A. In the description above L would be 5. Within the h^{th} stratum, multiple surface sediment samples are collected from n_h locations using an unbiased statistically valid sampling design. Surface PCB concentrations (x_{hi} $i=1,2,\ldots n_h$) are measured at each location. In what follows these values are assumed to be individual samples. In practice these may be composite samples although for simplicity, the following formulas assume single samples. The equations that follow could be applied to composite samples, or if discrete and composites are to be combined these equations can be modified slightly to accommodate composite sampling.

Further assume that there may be $h=1,2,\ldots M$ strata with area B_h that have not been sampled but which have been assumed to have known average surficial PCB concentrations (y_h, $h=1,2,\ldots,M$). Further assume that these strata have combined area given by the sum of the stratum areas B. In practice it is anticipated that there would be at most one stratum that would not have been sampled, but the more general case is illustrated here.

To estimate the overall SWAC for the collection of strata a weighted average of stratum averages is applied. Stratum means and sampling variances are first calculated as

$$\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}; \quad s_h^s = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(x_{hi} - \bar{x}_h\right)^2$$

These stratum specific estimates of the mean are combined across strata using the area weighted average

$$\bar{x}_{st} = \frac{\sum_{h=1}^{L}(A_h \times \bar{x}_h)}{\sum_{h=1}^{L} A_h} = \frac{\sum_{h=1}^{L}(A_h \times \bar{x}_h)}{A}$$

and the sampling variance of this weighted average is

$$\mathrm{var}(\bar{x}_{st}) = \frac{1}{A^2}\sum_{h=1}^{L} A_h^2 \times \frac{s_h^2}{n_h}$$

The M unsampled strata can be incorporated into the estimated average, however it would not be generally possible to estimate the precision of these areas due to the lack of sample data with which to estimate sample to sample variation and subsequently variance of the estimated mean. The stratified estimate of the unsampled areas is given by

$$\bar{y}_{st} = \frac{1}{B}\sum_{i=1}^{n_h}(B_h \times y_h).$$

Finally, the overall estimated SWAC is given by the weighted average of these two stratified sampling estimators

$$SWAC_{Estimate} = \frac{A \times \bar{x}_{st} + B \times \bar{y}_{st}}{A+B}.$$

Assuming that the variance of $\bar{y}_{st}$ is known or can be approximated the sampling variance of $SWAC_{estimate}$ is

$$\mathrm{var}(SWAC_{estimate}) = \frac{A^2\ \mathrm{var}(\bar{x}_{st}) + B^2\ \mathrm{var}(\bar{y}_{st})}{(A+B)^2}.$$

If the values in the unsampled areas are truly thought to be known, then the variance of $\bar{y}_{st}$ would be zero and the variance of $SWAC_{estimate}$ simplifies to

$$\mathrm{var}(SWAC_{estimate}) = \frac{A^2\ \mathrm{var}(\bar{x}_{st})}{(A+B)^2}.$$

Confidence Intervals

It is expected that each stratum will have relatively large numbers of confirmation samples. Because of these large sample sizes it is reasonable to estimate confidence limits based on the central limit theorem which states that for large sample sizes the mean is expected to have an approximately normal sampling distribution. Therefore approximate 100x(1-α)% confidence intervals are given by

$$SWAC_{estimate} \pm z_{1-\alpha/2} \times \sqrt{\mathrm{var}(SWAC_{estiamte})}$$

where $z_{1-\alpha/2}$ is a critical value of the standard normal distribution. For example for 95% confidence limits $\alpha = 0.05$ and $z_{1-\alpha/2} = 1.96$.

Discussion

If unsampled areas are small or negligible, then this estimate is dominated by the stratified sampling estimator of the sampled strata. However, if the unsampled areas are large relative to sampled strata, then the estimated *SWAC* will be dominated by the assumptions associated with the unsampled areas and its' sampling variance will reduce to essentially zero. For example if all sand covered areas are assumed to take on a particular concentration, the estimated *SWAC* would be only slightly different from the assumed value of the sand cover area and the confidence intervals would be artificially narrow. Uncertainty in this estimate is a combination of the sampling variation due to sampled strata and the bias associated with misspecification of assumptions in unsampled areas. The sensitivity to these assumptions can be determined by varying the assumed values y_h and plotting the range of confidence limits associated with the range of plausible assumptions. It is preferred that all stratum estimates are based on actual sample data from unbiased sampling designs so that estimates are unbiased and uncertainty is fully captured by the confidence limits.

References

Cochran, W.G. 1977. *Sampling Techniques*, *Third Edition*. John Wiley and Sons, New York.

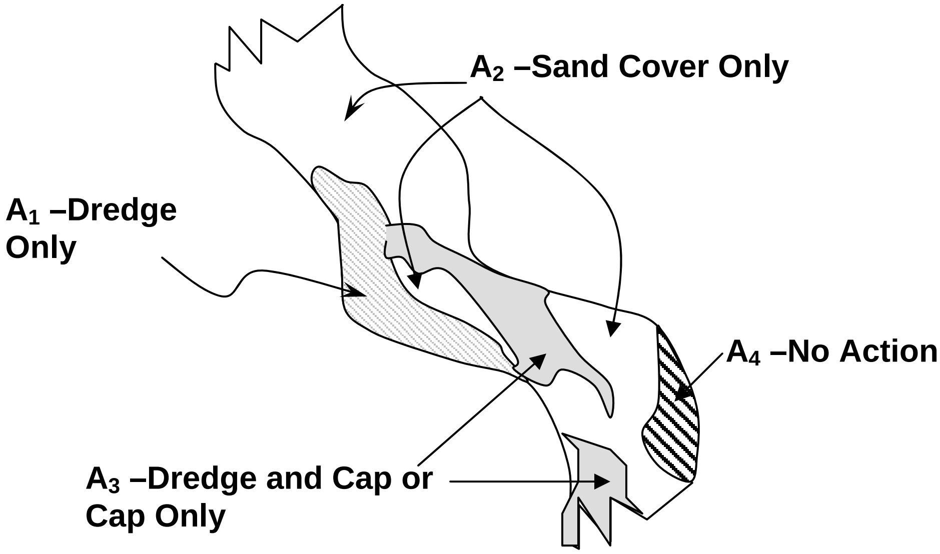

Figure 1. Schematic of strata associated with varying remedial activities.